CAAP Quarterly Report

Date of Report: *Jan 7, 2019*

Prepared for: *Thomas Finch (Project Manager) and Joshua Arnold (CAAP Program Manager), U.S. DOT Pipeline and Hazardous Materials Safety Administration*

Contract Number: *693JK31850005CAAP*

Project Title: *Low-variance Deep Graph Learning for Predictive Pipeline Assessment with Interacting Threats*

Prepared by: *Hao Zhang (Colorado School of Mines) and Yiming Deng (Michigan State University)*

Contact Information: *Dr. Hao Zhang, Department of Computer Science, Colorado School of Mines; 1500 Illinois St., Golden, CO 80401; Phone: 303-273-3581; Email: hzhang@mines.edu*

For quarterly period ending: *Jan 7, 2019*

Business and Activity Section

(a) Contract Activity

No contract modification was made or proposed in this quarterly period. No materials were purchased during this quarterly period.

(b) Status Update of Past Quarter Activities

Pipeline networks are essential to gather, transport, and distribute gas and hazardous liquid. Oil and gas revolution have been held back by aging pipeline infrastructure. Failures of pipeline infrastructure usually causes loss of life and safety issues that may cost millions of dollars in environmental cleanup, infrastructure repair, property damage, and production loss. Working towards achieving full energy independence by using both conventional and unconventional oil and natural gas to power our economy. The development and revival of pipeline infrastructure with seamless integration of computational algorithms and improved pipeline assessment techniques, and the advancement of nondestructive evaluation (NDE) inspection technologies and data analytics helps assure next generation pipeline system's usability, integrity, and safety.

A recent report has identified significant interacting threats and shown that, among many failures reported in the past few decades, Loss of Containment (LoC) may appear to have resulted from a single cause of an individual threat, but, in reality, has resulted because of circumstances that aggravated the threats [1]. However, characterization of coincident and overlapping defects, due to interacting threats has not been well addressed. For example, the interacting nature of the corrosion and mechanical damage consisting of a scrape or gouge is still debatable. Thus, a clear definition of coincident damage features and identification of interacting threats is a major existing technical gap. Another critical gap is that existing pipeline inspection techniques are designed for specific threat conditions and cannot provide a holistic assessment of interactive threats' impacts, which results in large sensor data variance and associated uncertainties, leading to significant difficulties in resolving overlapping damage, and thereby affecting integrity management decision adversely. A general framework for identification of coincident damage features with better quantification of material physical and geometric characteristics will be of critical importance for enhancing pipeline assessment methods and models in order to reduces variance.

In this reporting period, the research team performed comprehensive literature review, and made great progress toward achieving the technical objectives including:

(1) Designing methods for hybrid modeling of interacting threats and risk identification: We started designing and implementing a computational framework to understand NDE data characteristics for pipeline threats.
(2) Designing an approach for image data segmentation: We implemented a segmentation approach based on deep learning to segment objects from image data. The approach is currently evaluated using street view images and will be tested on pipeline threat data in the future work.
(3) Developing methods for characterization and diagnosis of interacting threats: We started developing low-variance characterization methods to characterize interacting threats using CNN-based deep learning methods.

The PIs believe that education is a critical component of the CAAP project, and we will integrate research with educational activities to prepare the next generation scientists and engineers for the gas and pipeline industry. In this reporting period, the research team made great progress toward the proposed educational objectives, including (1) involving three PhD students and several unpaid master and undergraduate students at Mines and MSU, (2) introducing the application of pipeline network inspection as an example in the courses (e.g., CSCI 473/573 Human-Centered Robotics at Mines) taught by the PIs, and (3) adapting the research topics from this project with the existing undergraduate research program (e.g., the Mines Undergraduate Research Honor Thesis) and MSU (e.g., ENSURE program).

(c) Cost Share Activity

PI Zhang used his 11.29% yearly effort as the in-kind cost share to work on the project at the Colorado School of Mines. Co-PI Yiming Deng used his 6.07% yearly effort as the in-kind cost share to work on the project at the Michigan State University. The cost share was used following the approved proposal and no modification was made.

(d) Performed Research: Developing and Evaluating New Methods for Low-Variance Interacting Threats Assessment

1. Background and Objectives in This Period of Performance

The key component in managing pipeline safety is threat identification. According to the ASME B31.8S standard [2], nine primary threat conditions are identified, which comes under three basic categories:

- Time-Dependent Threats (threats tending to grow over time)
    - Internal corrosion
    - External corrosion
    - Stress corrosion cracking
- Resident Threats (threats that do not grow over time; instead they tend to act when influenced by another condition or failure mechanism)
    - Manufacturing
    - Fabrication
    - Construction
- Time independent threats
    - Human error
    - Excavation damage
    - Earth movement, outside force or weather.

Pipeline operators use various methods and programs to prevent, detect, and mitigate and/or inspect for individual threats. Additionally, there may be circumstances when two or more threats can occur coincidentally and independently of each other. These “coincident threats” result in a likelihood of failure greater than that due to either threat individually or merely the superposition of the threats. Interactive threats are the merge of two or more defects in a pipe segment, the result of which is more damaging than either of the individual threat themselves. For instance, say after applying a neural network algorithm we are estimating the size and shape of the defects. We may predict that if the height of the anomaly crosses a certain threshold, it is fatal and requires immediate repair. Below that threshold the defect doesn’t need immediate repair. However, if there are several defects which are below that threshold and occur very close to each other then they can be combined as a large defect and can prove to be fatal. A matrix can be used to outline the relationship between different pairings of threats and detail the ways in which each particular active combination is managed. Not all boxes in the matrix would include guidance, as many combinations of threats are not inherently interactive. A simple representation of this method is demonstrated in Figure 1.

A matrix can also be applied to help quantify the synergistic effect in common threat interactions. Each operator’s approach to developing a matrix may be different based on historical threats and how they have been observed to potentially interact in different parts of its system. To capture interactive synergies, many operators use this type of matrix to apply “multiplier” factors to individual threat scores where the interaction is expected to exist. Another way of addressing potential threat interactions is to analyze and evaluate, generally through probabilistic or deterministic models, the coincident location of resident features (subcritical imperfections) that are acted upon by failure mechanisms, which can compound or magnify the original reduced resistance. This methodology couples the likelihood of a failure mechanism (or combination of failure mechanisms) being active and occurring at a resident feature location. A simple representation of this method is shown in Figure 2.

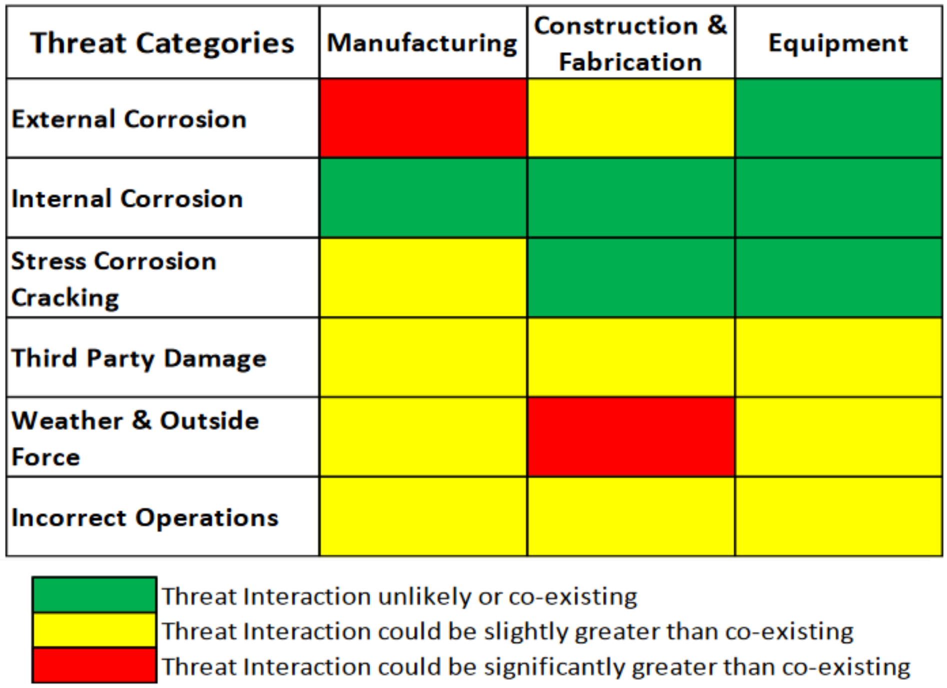

Figure 1: Different potential threat interactions

Due to significant uncertainty associated with different interactive damage types, complex pipe materials and external loadings, we propose a multimodal NDE sensing approach with capability to inspect multiple low variance interacting threats together with computational modelling for reliability evaluation and risk assessment of pipeline infrastructures. Additionally, pipeline integrity is time dependent and the associated uncertainty (i.e., damage types, severity) varies with time. Accurate uncertainty quantification and propagation analysis is critical for the validity of damage prognosis algorithms and health management approaches. Thus, an important requirement of the proposed system is to develop adaptive multimodal sensing and quantify its performance.

Several drawbacks of the current state-of-the-art (SOA) [3, 4, 5, 6, 7] are identified, including

- The existing SOA models cannot model the relationship of multiple threats from different categories.
- Existing models are based upon simple statistical models and thereby are prone to human error.
- Defects are growing continuous with time, and therefore effective data segmentation and threat prediction algorithm are required.
- Due to presence of noise and model uncertainty, it is challenging to estimate the size and shape of the defects.

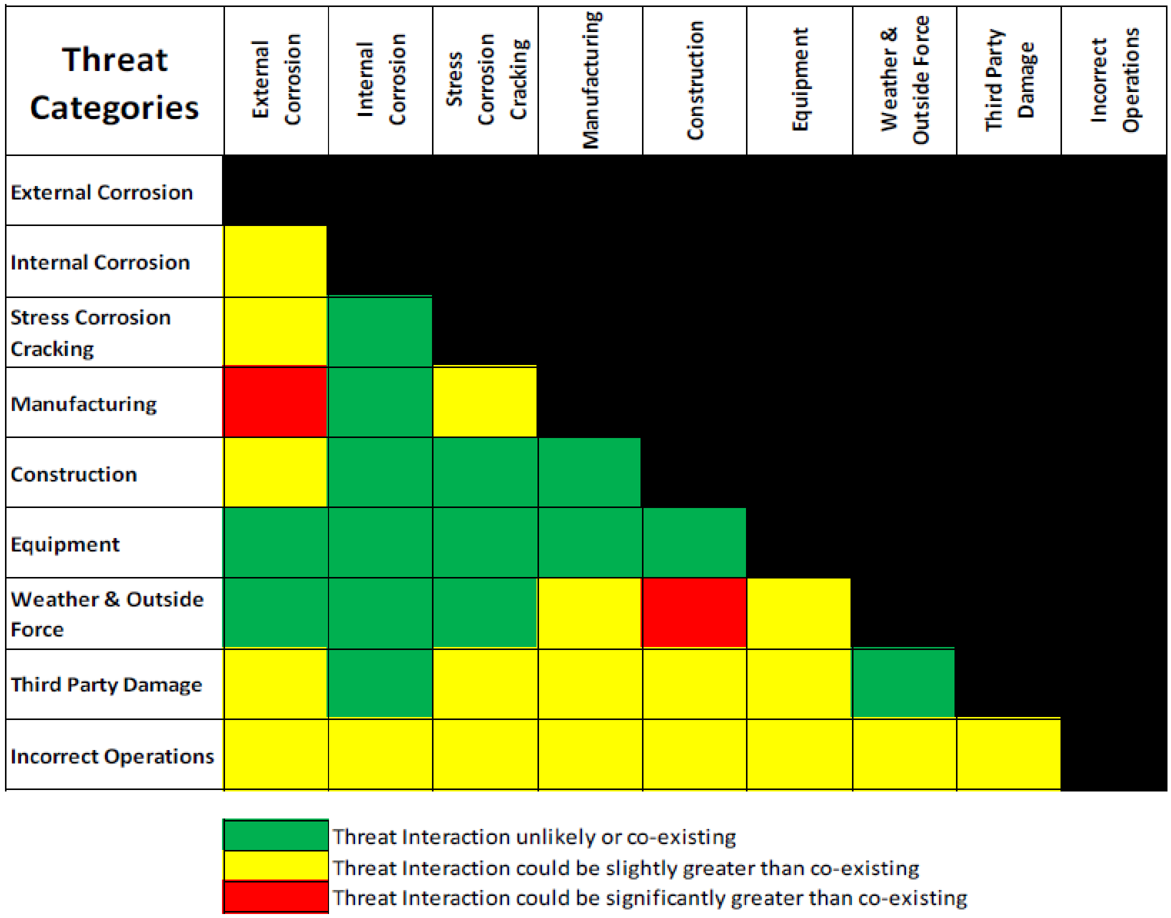

Figure 2: Interactive threat pairing

2. Preliminaries

This section describes the preliminaries needed to understand the performed research.

- **Mathematical Model**

  A mathematical model is a quantitative symbolic description of real-world systems. Mathematical models are approximations of the real-world. Here we only include the major factors into the models and ignore other factors. A good mathematical model can capture the underlying dynamics/rules of the real systems and provide predictions with sufficient accuracy. However, the goal is to construct an appropriate model but not an over-complicated or over-simplified one (as shown in Figure 3).

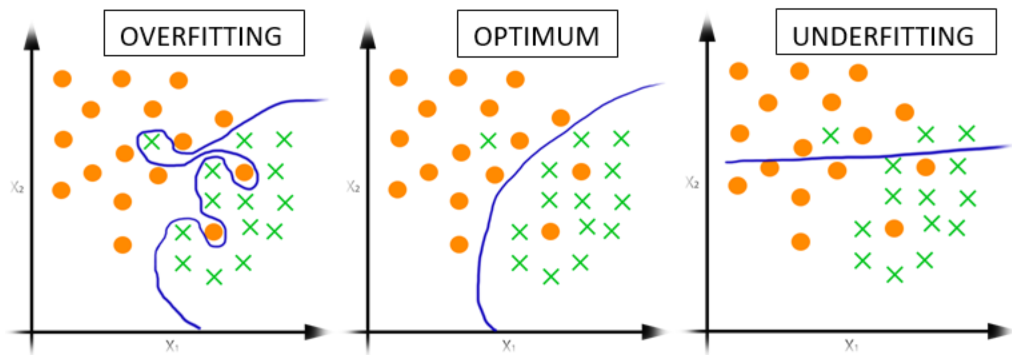

Figure 3: Illustration of overfitting, optimum, and under fitting models

- **Risk Identification**

  Research will focus on identifying risks and developing simulations (e.g., based on finite element models and statistical models) to generate heterogeneous In-line Inspection (ILI) Nondestructive Evaluation (NDE) data of individual and interacting threats.

- **CNN Based Data Characterization**

  Convoluted Neural Network (CNN) [8] is mostly applied to analyze visual imagery and it uses a variation of the multilayer perceptron to minimize feature extraction and preprocessing. CNN is not only a classifier but also a feature extractor. Here after FEM simulations of the defects by using different NDE techniques, we obtain perturbations in the magnetic field in X, Y, Z directions in the presence of anomalies. These deformations in magnetic fields are treated as images and the CNN is applied on these images for proper characterization and identification of future defects. However, the edge of the damage produces smaller deformation than the other damage parts and hence it is challenging to identify the edge of damages utilizing CNN. Thus, for better accuracy, long short-term memory (LSTM) is used as a secondary classifier. LSTM leverages the spatial structure of the defects in making estimates of the damage shape and volume and has been shown to greatly increase predictive capabilities of data driven systems.

- **Joint Learning**

  At first, we are performing anomaly detection using sensor output. Then, we are using supervised learning methods for defect characterization and reconstruction. For anomaly growth and prediction of unknown defects we are using the non-linear model CNN. For validation purpose we are dividing our data sets into two parts, the training data and test data. The steps of predicting defects are illustrated in Figure 4.

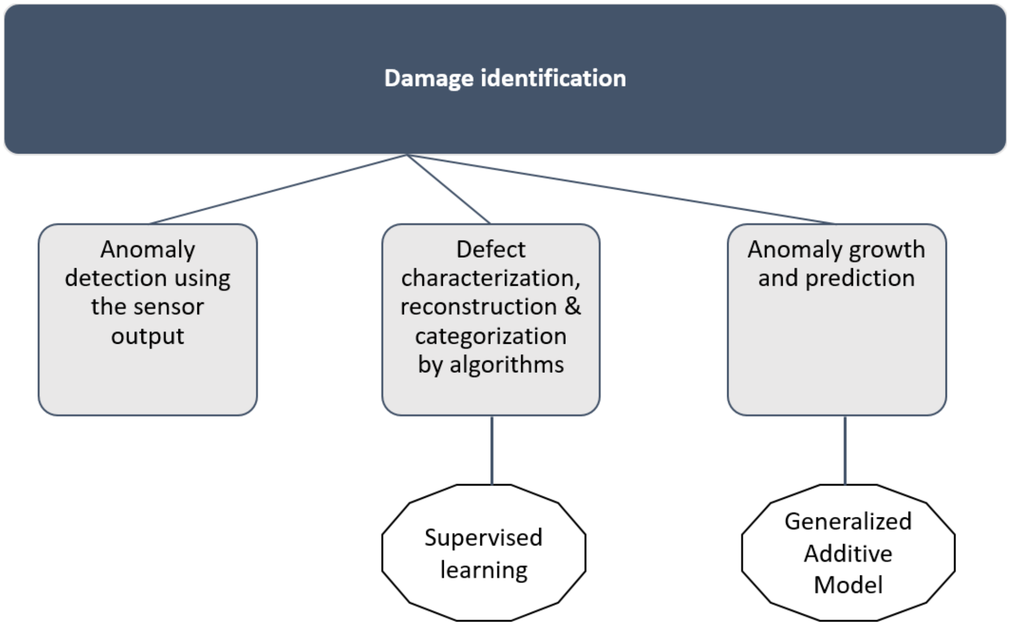

Figure 4: Steps involved in prediction of defects

- **Supervised Learning**

At first supervised learning techniques are to be applied for the identification and estimation of the size and shape of the defects. Due to complex geometry of the pipes it is not always possible to perform a visual inspection and moreover the direct inspections are time consuming. Next, with the aid of unsupervised learning algorithms, the shape of defects can be classified, and the defect size are estimated for new unknown defects. The initial introspection of supervised learning is as follows:

- Say: we have a set of known defect types. Types can be:
    1) Different sizes.
    2) Different shapes.
    3) Different location.
- We have inspection measurements from pipes with these defects.
- For a new pipe with unknown defect but having defect from one of the above categories how successfully we are recovering the defect category determines the efficiency of the algorithm.

The principle of supervised learning techniques is illustrated in Figure 5.

For the new unknown defects our goal is to detect the closest category to the new defect. This will enable us to approximate the unknown defect size, shape, and location. After extraction of the spatially adaptive features from the magnetic flux leakage (MFL) and pulsed eddy current (PEC) data we are feeding them into a classifier.

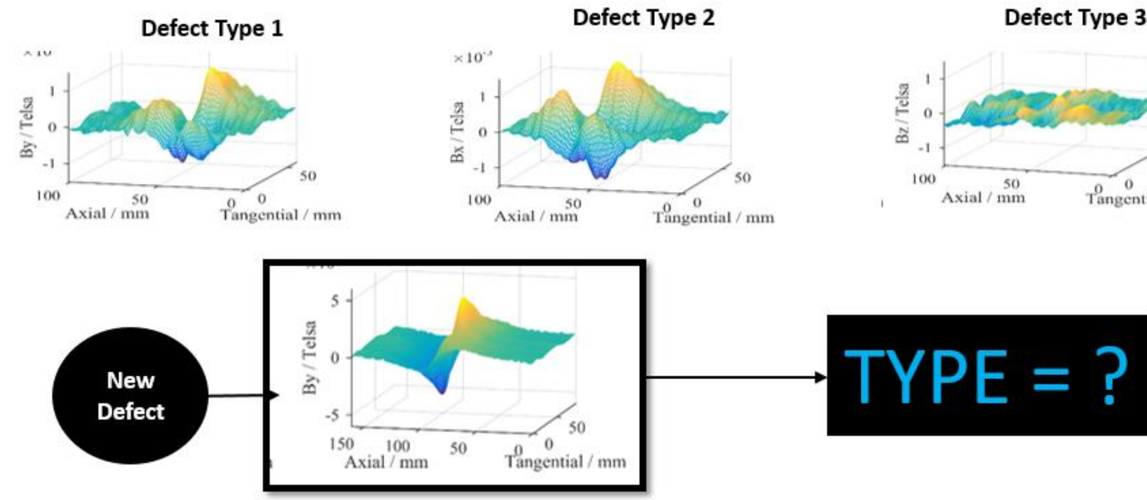

Figure 5: Working principle of supervised learning

The CNN model is constructed by one input layer, four convolutional layers, two pooling layers, one flatten layer, two dense layers, and a softmax function output layer. The convolutional layer can extract spatial features of every input image by the convolution operation, which operates on sliding windows of input. Here the images are treated as continuous video sequence which records the pipe condition continuously. The later frames are relevant with previous images. LSTM is a sequential model, which can learn from previous information by four gates: input gate, forget gate, output gate, and cell activation. In a sequence, the recurrent layer transfers the previous information (t-1) to the next (t) learning state.

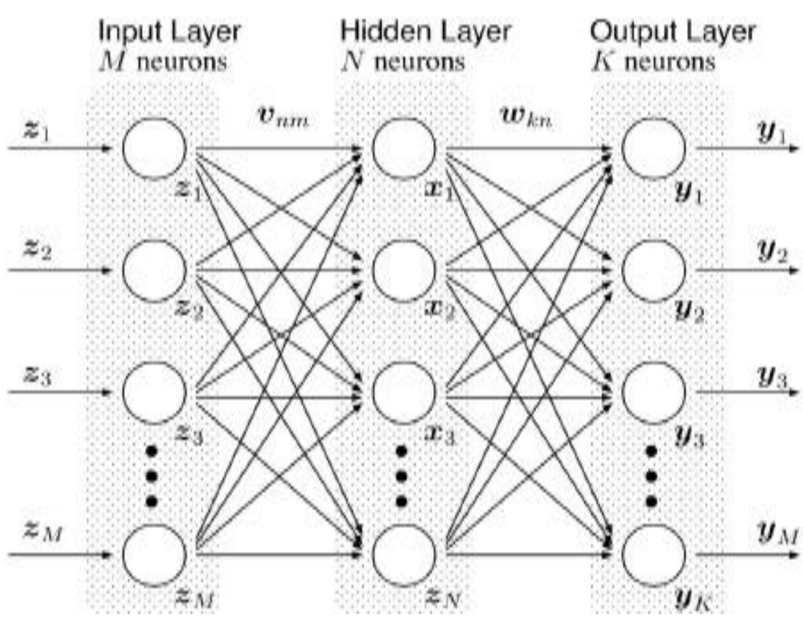

Figure 6: Basic Schematic of CNN

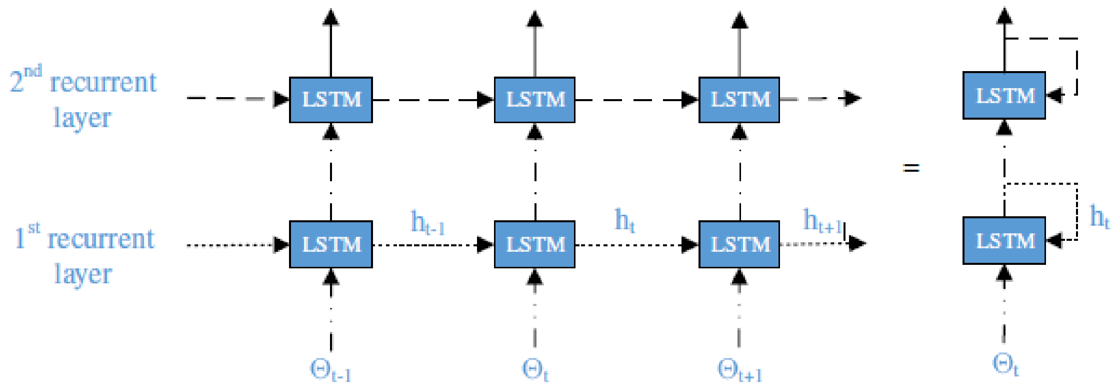

Figure 7: Basic Schematic of LSTM

3. Progress on Task 1: Hybrid Interacting Threat Modeling and Risk Identification

We will focus on developing computational models to generate and understand the heterogeneous In-line Inspection (ILI) data. Individual and interacting threats will be simulated by finite element methods (FEM) or statistical models. Finite element refers to the method from which the solution is numerically obtained from an arbitrary geometry by breaking it down into simple pieces called finite elements. Thus, a large amount of heterogeneous ILI NDE data of individual and interacting threats are produced. Individual and interacting defects are simulated based on FEM which is achieved via ANSYS Maxwell and COMSOL software. Modeling of finite element analysis supported by ANSYS and COMSOL is employed to study the relationship between geometric parameters and corresponding signals for pipeline defects. In ANSYS Maxwell 3D, the fundamental unit of the finite element is a tetrahedron. The Components of a Field that are tangential to the edges of an element are explicitly stored at the vertices; components of a field that is tangential to the face of an element and normal to an edge is explicitly stored at the midpoint of selected edges; The value of a vector field at an interior point is interpolated from the nodal values. To generate data, the key step is to first develop a theoretical model such as MFL, EC, or ultrasonic testing (UT) tool in motion inside the pipe. The model will adopt three-dimensional FEM of the ILI tool inside the pipe geometry. In MFL simulation, a three-dimensional model of cracks is built to explore the influence of MFL signal parameters including depth, width, inclination angle, and crack gap, etc. In the past, three-axis high-resolution MFL inspection can only detect the defects with a large opening. At present, much research work has been done in the field of signal analysis and quantitative analysis of three-axis high-resolution MFL. New MFL tool is designed to generate a field of excitation in the circumferential or transverse direction, thus being capable of detecting small axial cracks. In PEC, information about the nature, location and severity of a defect is obtained by subtracting the ‘no defect’, reference, or background signal from ‘defect’ signal.

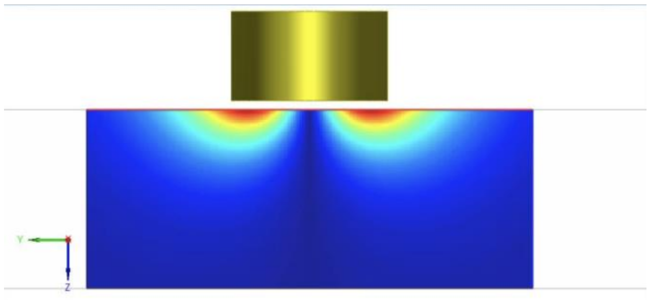

Figure 8: FEM modelling of PEC

The reference-subtracted signal shows peaks of different magnitudes and opposite polarity with the smaller one occurring later in time. The amplitudes of these peaks are affected by the change in position of the probe with respect to the defect position. FEM modeling has revealed that, for the probe dimensions used in the COMSOL, the maximum signal amplitude would be obtained when the outer surface of the pickup coil coincided with the center of the hole. In UT, FEM modeling provides physical insight into the nature of ultrasound/defect interactions and helps us to design data for planned ultrasonic tests. More importantly, when it is too difficult or expensive to conduct UT in the laboratory, simulation can achieve similar results as well as achieving benchmark studies for the validation of defect characterization scheme. ANSYS is a general-purpose finite element modeling package for numerically solving a wide variety of mechanical problems. These problems include: static/dynamic structural analysis (both linear and non-linear), heat transfer and fluid problems, as well as acoustic and electro-magnetic problems. In general, a finite element solution may be broken into the following three stages. This is a general guideline that can be used for setting up any finite element analysis.

- Preprocessing: defining the problem; the major steps in preprocessing are given below:
  - Define key points/lines/areas/volumes
  - Define element type and material/geometric properties
  - Mesh lines/areas/volumes as required
- The amount of detail required will depend on the dimensionality of the analysis (i.e. 1D, 2D, axis-symmetric, 3D).
- Solution: assigning loads, constraints and solving; here we specify the loads (point or pressure), constraints (translational and rotational) and finally solve the resulting set of equations.
- Postprocessing: further processing and viewing of the results; in this stage one may wish to see:
  - Lists of nodal displacements
  - Element forces and moments
  - Deflection plots
  - Stress contour diagrams

4. Progress on Task 2 – Image Segmentation for Extracting Interacting Threats

Identifying Regions of Interest (ROIs) is essential for threat characterization, since feature extraction and threat profiling are performed within the ROIs. The existing methods of ROI segmentation includes manual selection, thresholding (i.e., selecting pixels with values falling in a certain range to form a region) [9] and simple k-means clustering [10]. Pixel thresholding can result in errors and large segmentation noise, and manual and k-means methods cannot provide precise boundary and often result in big shape and size variance. Recently, deep learning, including convolutional neural networks (CNN) [8], has shown significant improvements on segmenting street views for self-driving. However, existing deep segmentation techniques are not directly applicable to solve the pipeline threat ROI extraction problem due to the following reasons. (1) Arbitrary threat sizes and shapes: Threats of the same type can show difference in size and shape; either big or small-sized threats can exceed the receptive field of a deep network and cause segmentation errors. (2) Co-occurrence of threats in interacting threats scenarios: Existing deep methods usually only consider single scale (e.g., at the pixel level) and lack the capability to incorporate the relationship of interacting threats for segmentation.

In this performance period, we started developing a deep network enhanced by a multiscale pooling mechanism to solve interacting threat ROI extraction and address the challenges. The proposed multiscale pooling introduces a hierarchy of pooling layers with each layer encoding information at a different scale. The coarsest level (at the top of the hierarchy) captures the relationship (i.e., co-occurrence) of the threats, and other levels capture the different sizes of threats, with a lower-level layer (more toward the bottom of the hierarchy) dealing with smaller sized threats. This multiscale pooling allows for modeling the co-occurrence of interacting threats and allows for modeling threats with different sizes. The multiscale pooling mechanism was integrated with CNN as a new component to address the challenges of pipeline threat segmentation.

The segmentation approach was implemented and initially tested on street view image data that was collected in the downtown of Golden near the campus of the Colorado School of Mines. We illustrate the preliminary results of the approach on street view images in Figure 9. It is observed a satisfactory segmentation was achieved, such that the city sign, sky, road, buildings, and vehicles are separated. Future research on this task will focus on adapting this approach to specifically address ROI extraction for interacting threats.

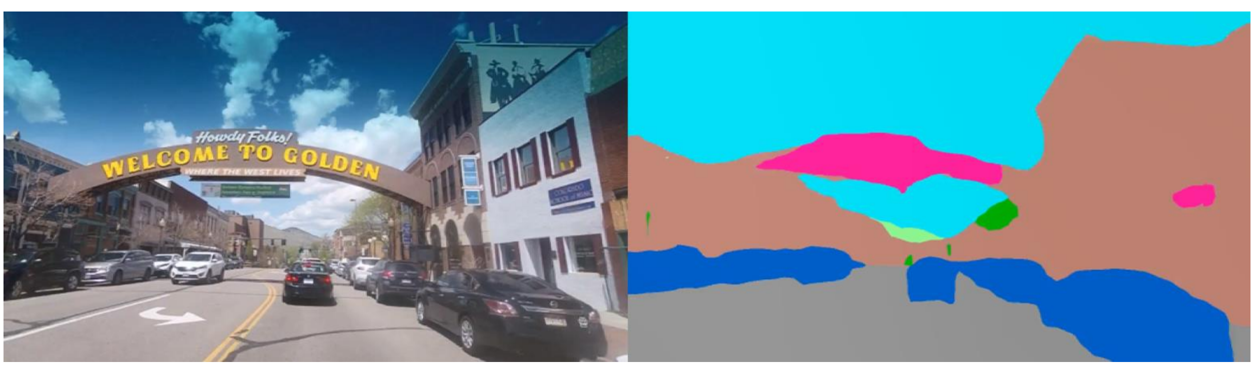

Figure 9: ROI segmentation results over street view images in preliminary experiments

5. Progress on Task 3 – Interacting Threats Characterization Assisted by Hybrid Models

This research will focus on characterizing individual and interacting threats by combining robust signal processing and characterization algorithms with a finite element model, and Convolutional

Neural Network (CNN)-based methods for threat diagnosis. Here by performing MFL and PEC simulations on ANSYS, we have witnessed the perturbation of magnetic field in the presence of different interactive defects and these defects are further characterized by machine learning algorithms. Specifically, CNN and supervised learning to estimate the location and severity of threats in the pipe. Here sophisticated threat characterization approaches are implemented that compensate for tool velocity, pipe grade, stress, pipe wall thickness and remnant magnetization, in order to provide a full profile of threats in the pipe wall. We will evaluate multiple candidate characterization approaches, based on neural network approach using CNN and LSTM and thereby selecting the method with lowest variance. To reduce the dimensionality of the data we will be using principal component analysis (PCA) which will serve as a feature extraction algorithm. We are dealing with images here but since the dimensionality of the data is high we are using PCA for feature extraction. Given a set of points in Euclidian space, the first principal component corresponds to a line that passes through the multidimensional mean and minimizes the sum of squares of the distances of the points from the line. The second principal component corresponds to the same concept after all correlation with the first principal component has been subtracted from the points. The singular values (in $\mathbf{\Sigma}$) are the square roots of the eigenvalues of the matrix $\mathbf{X}^{T}\mathbf{X}$. Each eigenvalue is proportional to the portion of the variance (more correctly of the sum of the squared distances of the points from their multidimensional mean) that is associated with each eigenvector. The sum of all the eigenvalues is equal to the sum of the squared distances of the points from their multidimensional mean. PCA essentially rotates the set of points around their mean to align with the principal components. This moves as much of the variance as possible (using an orthogonal transformation) into the first few dimensions. The values in the remaining dimensions, therefore, tend to be small and may be dropped with minimal loss of information. Hence PCA serves both data compression and invariance. In supervised learning, defect detection is basically a data classification problem to classify and characterize the defects. Here we have a known set of defect types, then for an unknown defect we have to predict in which category it will fall.

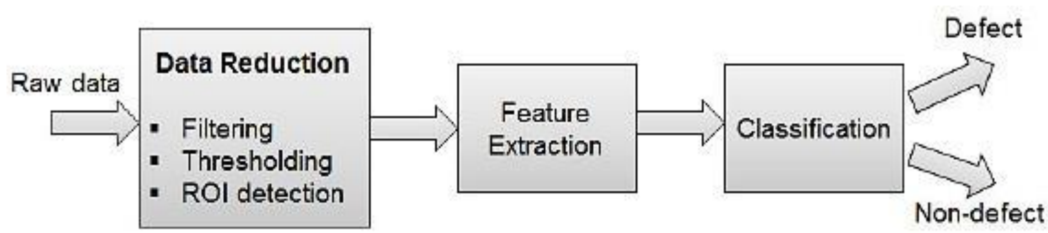

Figure 10: Framework of the proposed CNN-based automated threat diagnosis system

From the FEM simulations we have received the perturbation of magnetic field in the presence of defects like the above figure. From there, we know the location of the defects. To characterize the defects by their shapes and volumes we are passing these FEM figures as a collection of images and videos in CNN and LSTM (for edge detection) code developed in MATLAB, thereby predicting the accuracy of their characterization.

In simulation, material definition, boundary conditions, excitation, analysis, and solution setup should be thoroughly studied and predefined. For instance, boundary conditions that guide the behavior of the magnetic field at the interfaces or the edges of the problem region. It needs to be selected from three types of boundary: the natural boundary is adopted for the boundaries on the interface between objects. H field is continuous across the boundary; Neumann boundary condition is set up for exterior boundaries of solution domain. H field is tangential to the boundary and flux cannot cross it; The last one is the insulating layer condition, it is as Neumann except that current cannot cross the boundary. It is very useful to insulate two conductors which are in contact with each other. All the parameters must be considered so that calculations at each adaptive pass satisfy convergence criterion. This enables us to determine the permissible change in output quantity in percentile and evaluate output quantity at each adaptive pass. On the other hand, appropriate meshing in simulation is extremely important. The software uses the Finite Element Method (FEM) to solve Maxwell's equations. To obtain the set of algebraic equations to be solved, the geometry of the problem is discretized automatically into basic building blocks (e.g., tetrahedra in 3D). The assembly of all tetrahedra is referred to as the finite element mesh of the model or simply the mesh. Mesh plays an important role in accuracy of the computed results and thus requires higher mesh resolution in regions where field fields are of interest rapidly. Generally, mesh operation is implemented on all solids (model Objects) in the geometry automatically before the solution process is started. In Maxwell's Static Solvers, the mesh is automatically refined to achieve the required level of accuracy in field computation. This is referred as adaptive mesh refinement. For most of the cases, initial mesh is very coarse and close to uniform in size throughout the region. To achieve required level of accuracy in results, this mesh needs to be refined in areas where fields are of interest or the field gradients are high. Since adaptive meshing provides automated mesh refinement capability based on reported energy error in simulation, it is only available with static solvers. Therefore, to achieve higher accuracy, mesh operations shall be utilized.

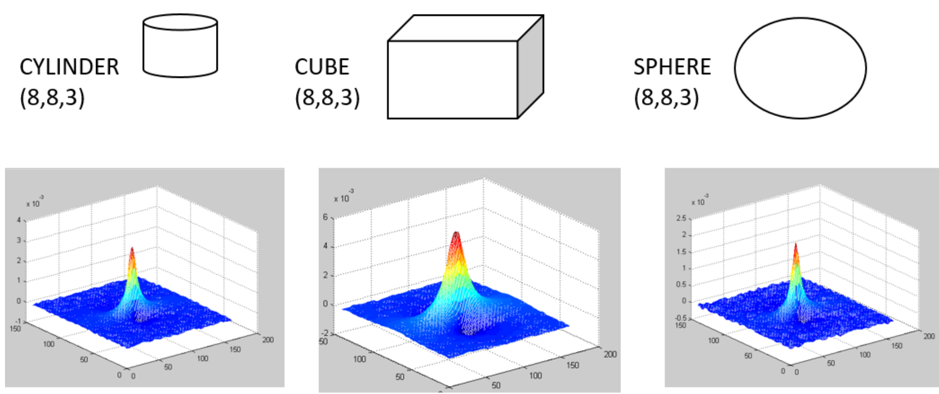

Figure 11: Characterization of defects

6. Summary and Future Work

In this period of performance, we reviewed the literature on methods to address interacting threats, and made great progress on developing methods for hybrid modeling of interacting threats and risk identification, deep learning based image data segmentation, and characterization and diagnosis of interacting threats. This research on predictive interacting threat assessment based on deep graph learning enables the missing capability to assess interacting threats with low variance. This research will result in reducing or avoiding the need for time-consuming manual model construction for interacting threat assessment, and establish a crucial body of knowledge of interacting threat properties and the knowledge needed to facilitate future design of interacting threat assessment models and standards.

In the next quarter, we continue making progress and completing the research tasks including heterogeneous modeling and deep learning based image segmentation, following the project schedule included in the approved proposal. We will also construct our mathematical models using the NDE data. Here we have used finite element modelling to simulate the defect models and then feed them to classification algorithms such as CNN and LSTM so that we can predict the shape and size of interactive threats with better accuracy. To promote education, we will continue involving PhD, Master's, and undergraduate students from our research group in the project, and advise them to improve their research skills as the project continues.

References

[1] H. A. Kishawy and H. A. Gabbar, " Review of pipeline integrity management practices," *International Journal of Pressure Vessels and Piping,* vol. 87, no. 7, pp. 373-380, 2010.

[2] "ASME B31.8S," [Online].

[3] A. Benjamin, J. Freire and R. Vieira, "Analysis of pipeline containing interacting corrosion defects," *Experimental Techniques,* vol. 31, no. 3, pp. 74-82, 2007.

[4] A. C. Benjamin, J. L. F. Freire, R. D. Vieira and D. J. Cunha, "Interaction of corrosion defects in pipelines–Part 1: Fundamentals," *International Journal of Pressure Vessels and Piping,* vol. 144, pp. 56-62, 2016.

[5] A. Benjamin, J. Freire, R. Vieira and D. Cunha, " Interaction of corrosion defects in pipelines–Part 2: MTI JIP database of corroded pipe tests," *International Journal of Pressure Vessels and Piping,* vol. 145, pp. 41-59, 2016.

[6] Interstate Natural Gas Association of America, "Interacting Threats to Pipeline Integrity – Defined and Explained," http://www.ingaa.org/File.aspx?id=20210, 2013.

[7] Northeast Gas Association, "NYSEARCH Pipeline Safety Update: Kiefner Interactive Threats Project," http://www.northeastgas.org/pdf/d_dzurko_safety.pdf, 2013.

[8] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE conference on computer vision and pattern recognition*, 2015.

[9] A. Sadr and R. S. Okhovat, "Extracting the region of interest from MFL signals," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 24, no. 2, pp. 427-434, 2016.

[10] S. Sangeetha, I. A. A. Daniel and R. Gokul, "Characterization of Defects in Magnetic Flux Leakage Images," *International Journal of Computer Applications,* 2013.